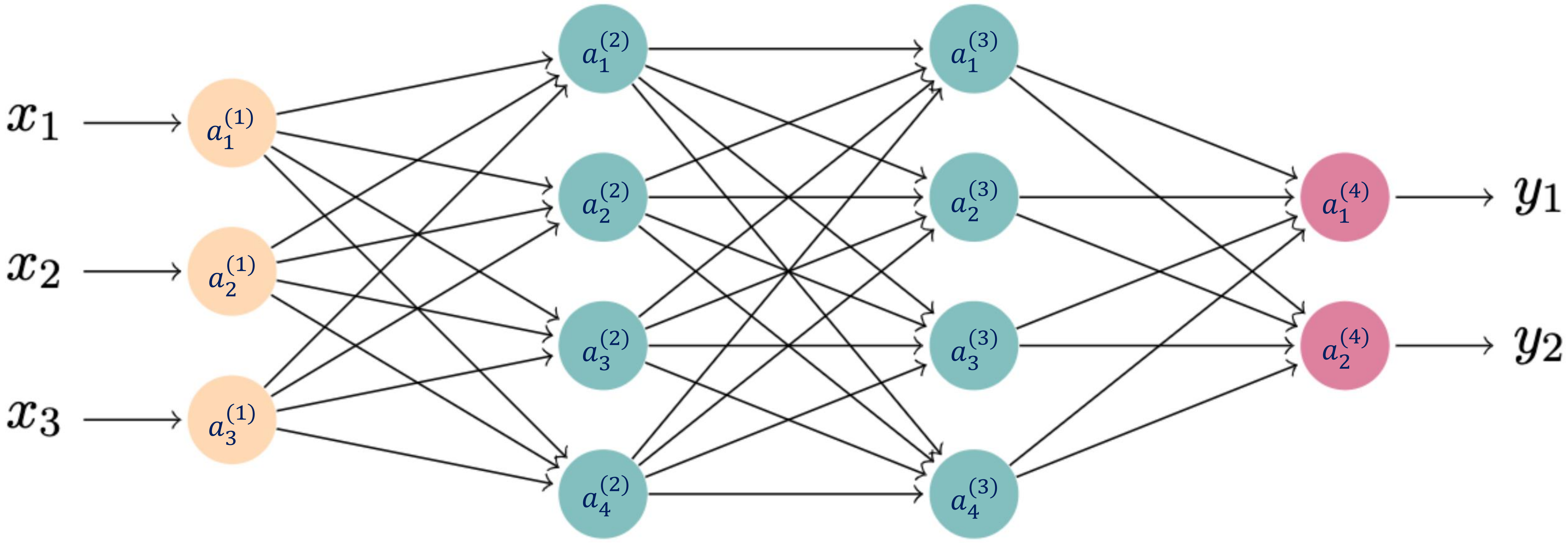
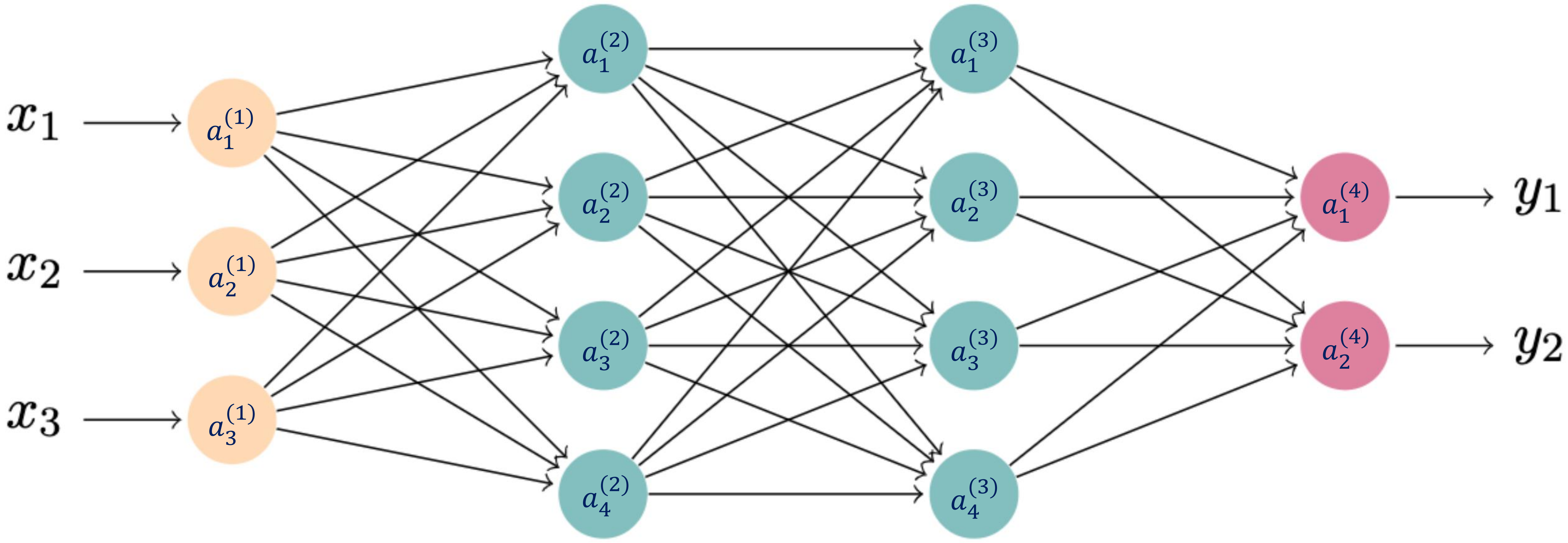
DEEP LEARNING
MACHINE LEARNING

By Suneesh Jacob


x_1
x_2
x_3
$a_1^{(1)}$
$a_2^{(1)}$
$a_3^{(1)}$
$a_1^{(2)}$
$a_2^{(2)}$
$a_3^{(2)}$
$a_4^{(2)}$
$a_1^{(3)}$
$a_2^{(3)}$
$a_3^{(3)}$
$a_4^{(3)}$
$a_1^{(4)}$
$a_2^{(4)}$
y_1
y_2



$$C = \frac{1}{2}\left[\frac{\left(a_1^{(4)} - y_1\right)^2 + \left(a_2^{(4)} - y_2\right)^2}{2}\right]$$

ACTIVATION VALUES, WEIGHTS AND BIASES

$$\begin{Bmatrix} a_1^{(4)} \\ a_2^{(4)} \end{Bmatrix} = \sigma\left(\begin{Bmatrix} z_1^{(4)} \\ z_2^{(4)} \end{Bmatrix}\right) = \begin{Bmatrix} \sigma\left(z_1^{(4)}\right) \\ \sigma\left(z_2^{(4)}\right) \end{Bmatrix}$$

$$\begin{Bmatrix} z_1^{(4)} \\ z_2^{(4)} \end{Bmatrix} = \begin{bmatrix} \mathrm{w}_{11}^{(3)} & \mathrm{w}_{12}^{(3)} & \mathrm{w}_{13}^{(3)} & \mathrm{w}_{14}^{(3)} \\ \mathrm{w}_{21}^{(3)} & \mathrm{w}_{22}^{(3)} & \mathrm{w}_{23}^{(3)} & \mathrm{w}_{24}^{(3)} \end{bmatrix} \begin{Bmatrix} a_1^{(3)} \\ a_2^{(3)} \\ a_3^{(3)} \\ a_4^{(3)} \end{Bmatrix} + \begin{Bmatrix} b_1^{(3)} \\ b_2^{(3)} \end{Bmatrix}$$

$$\begin{Bmatrix} a_1^{(3)} \\ a_2^{(3)} \\ a_3^{(3)} \\ a_4^{(3)} \end{Bmatrix} = \sigma\left(\begin{Bmatrix} z_1^{(3)} \\ z_2^{(3)} \\ z_3^{(3)} \\ z_4^{(3)} \end{Bmatrix}\right) = \begin{Bmatrix} \sigma\left(z_1^{(3)}\right) \\ \sigma\left(z_2^{(3)}\right) \\ \sigma\left(z_3^{(3)}\right) \\ \sigma\left(z_4^{(3)}\right) \end{Bmatrix}$$

$$\begin{Bmatrix} z_1^{(3)} \\ z_2^{(3)} \\ z_3^{(3)} \\ z_4^{(3)} \end{Bmatrix} = \begin{bmatrix} \mathrm{w}_{11}^{(2)} & \mathrm{w}_{12}^{(2)} & \mathrm{w}_{13}^{(2)} & \mathrm{w}_{14}^{(2)} \\ \mathrm{w}_{21}^{(2)} & \mathrm{w}_{22}^{(2)} & \mathrm{w}_{23}^{(2)} & \mathrm{w}_{24}^{(2)} \\ \mathrm{w}_{31}^{(2)} & \mathrm{w}_{32}^{(2)} & \mathrm{w}_{33}^{(2)} & \mathrm{w}_{34}^{(2)} \\ \mathrm{w}_{41}^{(2)} & \mathrm{w}_{42}^{(2)} & \mathrm{w}_{43}^{(2)} & \mathrm{w}_{44}^{(2)} \end{bmatrix} \begin{Bmatrix} a_1^{(2)} \\ a_2^{(2)} \\ a_3^{(2)} \\ a_4^{(2)} \end{Bmatrix} + \begin{Bmatrix} b_1^{(2)} \\ b_2^{(2)} \\ b_3^{(2)} \\ b_4^{(2)} \end{Bmatrix}$$

$$\begin{Bmatrix} a_1^{(2)} \\ a_2^{(2)} \\ a_3^{(2)} \\ a_4^{(2)} \end{Bmatrix} = \sigma\left(\begin{Bmatrix} z_1^{(2)} \\ z_2^{(2)} \\ z_3^{(2)} \\ z_4^{(2)} \end{Bmatrix}\right) = \begin{Bmatrix} \sigma\left(z_1^{(2)}\right) \\ \sigma\left(z_2^{(2)}\right) \\ \sigma\left(z_3^{(2)}\right) \\ \sigma\left(z_4^{(2)}\right) \end{Bmatrix}$$

$$\begin{Bmatrix} z_1^{(2)} \\ z_2^{(2)} \\ z_3^{(2)} \\ z_4^{(2)} \end{Bmatrix} = \begin{bmatrix} \mathrm{w}_{11}^{(1)} & \mathrm{w}_{12}^{(1)} & \mathrm{w}_{13}^{(1)} \\ \mathrm{w}_{21}^{(1)} & \mathrm{w}_{22}^{(1)} & \mathrm{w}_{23}^{(1)} \\ \mathrm{w}_{31}^{(1)} & \mathrm{w}_{32}^{(1)} & \mathrm{w}_{33}^{(1)} \\ \mathrm{w}_{41}^{(1)} & \mathrm{w}_{42}^{(1)} & \mathrm{w}_{43}^{(1)} \end{bmatrix} \begin{Bmatrix} a_1^{(1)} \\ a_2^{(1)} \\ a_3^{(1)} \end{Bmatrix} + \begin{Bmatrix} b_1^{(1)} \\ b_2^{(1)} \\ b_3^{(1)} \\ b_4^{(1)} \end{Bmatrix}$$

$$\sigma(x) = \frac{1}{1 + e^{-x}}$$

$$\begin{aligned}\sigma'(x) &= \frac{\partial}{\partial x}\left(\sigma(x)\right) \\ &= \frac{-1}{(1 + e^{-x})^2} \times -e^{-x} \\ &= \frac{e^{-x}}{(1 + e^{-x})^2} \\ &= \frac{1}{1 + e^{-x}} \times \frac{1 + e^{-x} - 1}{1 + e^{-x}} \\ &= \frac{1}{1 + e^{-x}} \times \left(1 - \frac{1}{1 + e^{-x}}\right) \\ &= \sigma(x)\left(1 - \sigma(x)\right)\end{aligned}$$

$$z_k^{(L+1)} = \sum_{j=1}^{n_L} w_{kj}^{(L)} a_j^{(L)} + b_k^{(L)}$$

$$a_k^{(L+1)} = \sigma\left(z_k^{(L+1)}\right)$$

$$\delta_j^{(L)} = \frac{\partial C}{\partial a_j^{(L)}}$$

$$\delta_j^{(L)} = \sum_{k=1}^{n_L} \delta_k^{(L+1)} \frac{\partial a_k^{(L+1)}}{\partial a_j^{(L)}}$$

$$\frac{\partial a_k^{(L+1)}}{\partial a_j^{(L)}} = \frac{\partial \sigma\left(z_k^{(L+1)}\right)}{\partial z_k^{(L+1)}} \frac{\partial z_k^{(L+1)}}{\partial a_j^{(L)}} = \sigma'\left(z_k^{(L+1)}\right) w_{kj}^{(L)}$$

$$\sigma'\left(z_k^{(L+1)}\right) \frac{\partial \sigma\left(z_k^{(L+1)}\right)}{\partial a_k^{(L+1)}} = \left[\sigma\left(z_k^{(L+1)}\right)\right]\left[1 - \sigma\left(z_k^{(L+1)}\right)\right]$$

DELTA VECTORS

$$\delta^{(4)} = \begin{Bmatrix} \delta_1^{(4)} \\ \delta_2^{(4)} \end{Bmatrix} = \begin{Bmatrix} \dfrac{\partial C}{\partial a_1^{(4)}} \\ \dfrac{\partial C}{\partial a_2^{(4)}} \end{Bmatrix}$$

$$\delta^{(3)} = \begin{Bmatrix} \delta_1^{(3)} \\ \delta_2^{(3)} \\ \delta_3^{(3)} \\ \delta_4^{(3)} \end{Bmatrix} = \begin{Bmatrix} \dfrac{\partial C}{\partial a_1^{(3)}} \\ \dfrac{\partial C}{\partial a_2^{(3)}} \\ \dfrac{\partial C}{\partial a_3^{(3)}} \\ \dfrac{\partial C}{\partial a_4^{(3)}} \end{Bmatrix}$$

$$\delta^{(2)} = \begin{Bmatrix} \delta_1^{(2)} \\ \delta_2^{(2)} \\ \delta_3^{(2)} \\ \delta_4^{(2)} \end{Bmatrix} = \begin{Bmatrix} \dfrac{\partial C}{\partial a_1^{(2)}} \\ \dfrac{\partial C}{\partial a_2^{(2)}} \\ \dfrac{\partial C}{\partial a_3^{(2)}} \\ \dfrac{\partial C}{\partial a_4^{(2)}} \end{Bmatrix}$$

$$\delta^{(1)} = \begin{Bmatrix} \delta_1^{(1)} \\ \delta_2^{(1)} \\ \delta_3^{(1)} \end{Bmatrix} = \begin{Bmatrix} \dfrac{\partial C}{\partial a_1^{(4)}} \\ \dfrac{\partial C}{\partial a_1^{(4)}} \\ \dfrac{\partial C}{\partial a_1^{(4)}} \end{Bmatrix}$$

GRADIENT

$$C = \frac{1}{2}\left[\frac{\left(a_1^{(4)} - y_1\right)^2 + \left(a_2^{(4)} - y_2\right)^2}{2}\right]$$

$$\begin{Bmatrix} \delta_1^{(4)} \\ \delta_2^{(4)} \end{Bmatrix} = \begin{Bmatrix} \frac{\partial C}{\partial a_1^{(4)}} \\ \frac{\partial C}{\partial a_2^{(4)}} \end{Bmatrix} = \begin{Bmatrix} \frac{\left(a_1^{(4)} - y_1\right)}{2} \\ \frac{\left(a_2^{(4)} - y_2\right)}{2} \end{Bmatrix}$$

$$\begin{Bmatrix} \delta_1^{(3)} \\ \delta_2^{(3)} \\ \delta_3^{(3)} \\ \delta_4^{(3)} \end{Bmatrix} = \begin{bmatrix} w_{11}^{(3)} & w_{21}^{(3)} \\ w_{12}^{(3)} & w_{22}^{(3)} \\ w_{13}^{(3)} & w_{23}^{(3)} \\ w_{14}^{(3)} & w_{24}^{(3)} \end{bmatrix} \begin{Bmatrix} \delta_1^{(4)}\sigma'\left(z_1^{(4)}\right) \\ \delta_2^{(4)}\sigma'\left(z_2^{(4)}\right) \end{Bmatrix}$$

$$\begin{Bmatrix} \delta_1^{(2)} \\ \delta_2^{(2)} \\ \delta_3^{(2)} \\ \delta_4^{(2)} \end{Bmatrix} = \begin{bmatrix} w_{11}^{(2)} & w_{21}^{(2)} & w_{31}^{(2)} & w_{41}^{(2)} \\ w_{12}^{(2)} & w_{22}^{(2)} & w_{32}^{(2)} & w_{42}^{(2)} \\ w_{13}^{(2)} & w_{23}^{(2)} & w_{33}^{(2)} & w_{43}^{(2)} \\ w_{14}^{(2)} & w_{24}^{(2)} & w_{34}^{(2)} & w_{44}^{(2)} \end{bmatrix} \begin{Bmatrix} \delta_1^{(3)}\sigma'\left(z_1^{(3)}\right) \\ \delta_2^{(3)}\sigma'\left(z_2^{(3)}\right) \\ \delta_3^{(3)}\sigma'\left(z_3^{(3)}\right) \\ \delta_4^{(3)}\sigma'\left(z_4^{(3)}\right) \end{Bmatrix}$$

$$\begin{Bmatrix} \delta_1^{(1)} \\ \delta_2^{(1)} \\ \delta_3^{(1)} \end{Bmatrix} = \begin{bmatrix} w_{11}^{(1)} & w_{21}^{(1)} & w_{31}^{(1)} & w_{41}^{(1)} \\ w_{12}^{(1)} & w_{22}^{(1)} & w_{32}^{(1)} & w_{42}^{(1)} \\ w_{13}^{(1)} & w_{23}^{(1)} & w_{33}^{(1)} & w_{43}^{(1)} \end{bmatrix} \begin{Bmatrix} \delta_1^{(2)}\sigma'\left(z_1^{(2)}\right) \\ \delta_2^{(2)}\sigma'\left(z_2^{(2)}\right) \\ \delta_3^{(2)}\sigma'\left(z_3^{(2)}\right) \\ \delta_4^{(2)}\sigma'\left(z_4^{(2)}\right) \end{Bmatrix}$$

GRADIENT

$$\frac{\partial C}{\partial w_{ij}^{(L)}} = \frac{\partial C}{\partial a_i^{(L+1)}} \frac{\partial a_i^{(L+1)}}{\partial w_{ij}^{(L)}}$$

$$\frac{\partial C}{\partial w_{ij}^{(L)}} = \delta_i^{(L+1)} \sigma'\left(z_i^{(L+1)}\right) a_i^{(L)}$$

$$\frac{\partial C}{\partial b_i^{(L)}} = \delta_i^{(L+1)} \sigma'\left(z_i^{(L+1)}\right)$$

$$\left\{\frac{\partial C}{\partial b^{(3)}}\right\} = \begin{Bmatrix} \delta_1^{(4)} \sigma'\left(z_1^{(4)}\right) \\ \delta_2^{(4)} \sigma'\left(z_2^{(4)}\right) \end{Bmatrix}$$

$$\left\{\frac{\partial C}{\partial b^{(2)}}\right\} = \begin{Bmatrix} \delta_1^{(3)} \sigma'\left(z_1^{(3)}\right) \\ \delta_2^{(3)} \sigma'\left(z_2^{(3)}\right) \\ \delta_3^{(3)} \sigma'\left(z_3^{(3)}\right) \\ \delta_4^{(3)} \sigma'\left(z_4^{(3)}\right) \end{Bmatrix}$$

$$\left\{\frac{\partial C}{\partial b^{(1)}}\right\} = \begin{Bmatrix} \delta_1^{(2)} \sigma'\left(z_1^{(2)}\right) \\ \delta_2^{(2)} \sigma'\left(z_2^{(2)}\right) \\ \delta_3^{(2)} \sigma'\left(z_3^{(2)}\right) \\ \delta_4^{(2)} \sigma'\left(z_4^{(2)}\right) \end{Bmatrix}$$

GRADIENT

$$\frac{\partial C}{\partial w_{ij}^{(L)}} = \frac{\partial C}{\partial a_i^{(L+1)}} \frac{\partial a_i^{(L+1)}}{\partial w_{ij}^{(L)}}$$

$$\frac{\partial C}{\partial w_{ij}^{(L)}} = \delta_i^{(L+1)} \sigma'\left(z_i^{(L+1)}\right) a_i^{(L)}$$

$$\frac{\partial C}{\partial b_i^{(L)}} = \delta_i^{(L+1)} \sigma'\left(z_i^{(L+1)}\right)$$

$$\left[\frac{\partial C}{\partial w^{(3)}}\right] = \begin{bmatrix} \frac{\partial C}{\partial b_1^{(3)}} a_1^{(3)} & \frac{\partial C}{\partial b_1^{(3)}} a_2^{(3)} & \frac{\partial C}{\partial b_1^{(3)}} a_3^{(3)} & \frac{\partial C}{\partial b_1^{(3)}} a_4^{(3)} \\ \frac{\partial C}{\partial b_2^{(3)}} a_1^{(3)} & \frac{\partial C}{\partial b_2^{(3)}} a_2^{(3)} & \frac{\partial C}{\partial b_2^{(3)}} a_3^{(3)} & \frac{\partial C}{\partial b_2^{(3)}} a_4^{(3)} \end{bmatrix} = \left\{\frac{\partial C}{\partial b^{(3)}}\right\} \left\{a^{(3)}\right\}^T$$

$$\left[\frac{\partial C}{\partial w^{(2)}}\right] = \begin{bmatrix} \frac{\partial C}{\partial b_1^{(2)}} a_1^{(2)} & \frac{\partial C}{\partial b_1^{(2)}} a_2^{(2)} & \frac{\partial C}{\partial b_1^{(2)}} a_3^{(2)} & \frac{\partial C}{\partial b_1^{(2)}} a_4^{(2)} \\ \frac{\partial C}{\partial b_2^{(2)}} a_1^{(2)} & \frac{\partial C}{\partial b_2^{(2)}} a_2^{(2)} & \frac{\partial C}{\partial b_2^{(2)}} a_3^{(2)} & \frac{\partial C}{\partial b_2^{(2)}} a_4^{(2)} \\ \frac{\partial C}{\partial b_3^{(2)}} a_1^{(2)} & \frac{\partial C}{\partial b_3^{(2)}} a_2^{(2)} & \frac{\partial C}{\partial b_3^{(2)}} a_3^{(2)} & \frac{\partial C}{\partial b_3^{(2)}} a_4^{(2)} \\ \frac{\partial C}{\partial b_4^{(2)}} a_1^{(2)} & \frac{\partial C}{\partial b_4^{(2)}} a_2^{(2)} & \frac{\partial C}{\partial b_4^{(2)}} a_3^{(2)} & \frac{\partial C}{\partial b_4^{(2)}} a_4^{(2)} \end{bmatrix} = \left\{\frac{\partial C}{\partial b^{(2)}}\right\} \left\{a^{(2)}\right\}^T$$

$$\left[\frac{\partial C}{\partial w^{(1)}}\right] = \begin{bmatrix} \frac{\partial C}{\partial b_1^{(1)}} a_1^{(1)} & \frac{\partial C}{\partial b_1^{(1)}} a_2^{(1)} & \frac{\partial C}{\partial b_1^{(1)}} a_3^{(1)} \\ \frac{\partial C}{\partial b_2^{(1)}} a_1^{(1)} & \frac{\partial C}{\partial b_2^{(1)}} a_2^{(1)} & \frac{\partial C}{\partial b_2^{(1)}} a_3^{(1)} \\ \frac{\partial C}{\partial b_3^{(1)}} a_1^{(1)} & \frac{\partial C}{\partial b_3^{(1)}} a_2^{(1)} & \frac{\partial C}{\partial b_3^{(1)}} a_3^{(1)} \\ \frac{\partial C}{\partial b_4^{(1)}} a_1^{(1)} & \frac{\partial C}{\partial b_4^{(1)}} a_2^{(1)} & \frac{\partial C}{\partial b_4^{(1)}} a_3^{(1)} \end{bmatrix} = \left\{\frac{\partial C}{\partial b^{(1)}}\right\} \left\{a^{(1)}\right\}^T$$

BACK PROPAGATION FORMULAE

$$\left\{\frac{\partial C}{\partial b^{(L)}}\right\} = \left\{\delta^{(L+1)}\right\} \odot \left\{\sigma'\left(z^{(L+1)}\right)\right\}$$

$$\left\{\delta^{(L)}\right\} = \frac{\partial C}{\partial a^{(L)}} = \left[w^{(L)}\right]^T \left\{\frac{\partial C}{\partial b^{(L)}}\right\}$$

$$\left[\frac{\partial C}{\partial w^{(L)}}\right] = \left\{\frac{\partial C}{\partial b^{(L)}}\right\} \left\{a^{(L)}\right\}^T$$

$$\left\{\frac{\partial C}{\partial b^{(3)}}\right\} = \left\{\delta^{(4)}\right\} \odot \left\{\sigma'\left(z^{(4)}\right)\right\} = \begin{Bmatrix} \delta_1^{(4)}\sigma'\left(z_1^{(4)}\right) \\ \delta_2^{(4)}\sigma'\left(z_2^{(4)}\right) \end{Bmatrix}$$

$$\left[\frac{\partial C}{\partial w^{(3)}}\right] = \left\{\frac{\partial C}{\partial b^{(3)}}\right\} \left\{a^{(3)}\right\}^T$$

$$\left\{\delta^{(3)}\right\} = \left[w^{(3)}\right]^T \left\{\frac{\partial C}{\partial b^{(3)}}\right\}$$

$$\left\{\frac{\partial C}{\partial b^{(2)}}\right\} = \left\{\delta^{(3)}\right\} \odot \left\{\sigma'\left(z^{(3)}\right)\right\}$$

$$\left[\frac{\partial C}{\partial w^{(2)}}\right] = \left\{\frac{\partial C}{\partial b^{(2)}}\right\} \left\{a^{(2)}\right\}^T$$

$$\left\{\delta^{(2)}\right\} = \left[w^{(2)}\right]^T \left\{\frac{\partial C}{\partial b^{(2)}}\right\}$$

$$\left\{\frac{\partial C}{\partial b^{(1)}}\right\} = \left\{\delta^{(2)}\right\} \odot \left\{\sigma'\left(z^{(2)}\right)\right\}$$

$$\left[\frac{\partial C}{\partial w^{(1)}}\right] = \left\{\frac{\partial C}{\partial b^{(1)}}\right\} \left\{a^{(1)}\right\}^T$$

PROBLEM

$$w^{(1)} = \frac{1}{10}\begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 7 & 8 & 9 \\ 10 & 11 & 12 \end{bmatrix}, b^{(1)} = \frac{1}{10}\begin{Bmatrix} 1 \\ 2 \\ 3 \\ 4 \end{Bmatrix}$$

$$w^{(2)} = \frac{1}{10}\begin{bmatrix} 1 & 2 & 3 & 4 \\ 5 & 6 & 7 & 8 \\ 9 & 10 & 11 & 12 \\ 13 & 14 & 15 & 16 \end{bmatrix}, b^{(2)} = \frac{1}{10}\begin{Bmatrix} 1 \\ 2 \\ 3 \\ 4 \end{Bmatrix}$$

$$w^{(3)} = \frac{1}{10}\begin{bmatrix} 1 & 2 & 3 & 4 \\ 5 & 6 & 7 & 8 \end{bmatrix}, b^{(3)} = \frac{1}{10}\begin{Bmatrix} 1 \\ 2 \end{Bmatrix}$$

$$x = \frac{1}{2}\begin{Bmatrix} 1 \\ 2 \\ 3 \end{Bmatrix}$$

$$y = \begin{Bmatrix} 5 \\ 10 \end{Bmatrix}$$

With these values, find the gradient of the loss function with respect to weights and biases, using back propagation.

SOLUTION

$$a^{(1)} = x = \frac{1}{2}\begin{Bmatrix} 1 \\ 2 \\ 3 \end{Bmatrix} = \begin{Bmatrix} 0.5 \\ 1 \\ 1.5 \end{Bmatrix}$$

$$z^{(2)} = w^{(1)}a^{(1)} + b^{(1)} = \begin{Bmatrix} 0.8 \\ 1.8 \\ 2.8 \\ 3.8 \end{Bmatrix}$$

$$a^{(2)} = \sigma\left(z^{(2)}\right) = \begin{Bmatrix} 0.69 \\ 0.86 \\ 0.94 \\ 0.98 \end{Bmatrix}$$

$$z^{(3)} = w^{(2)}a^{(2)} + b^{(2)} = \begin{Bmatrix} 1.01 \\ 2.5 \\ 3.99 \\ 5.48 \end{Bmatrix}$$

$$a^{(3)} = \sigma\left(z^{(3)}\right) = \begin{Bmatrix} 0.72 \\ 0.92 \\ 0.98 \\ 1 \end{Bmatrix}$$

$$z^{(4)} = w^{(3)}a^{(3)} + b^{(3)} = \begin{Bmatrix} 1.05 \\ 2.61 \end{Bmatrix}$$

$$a^{(4)} = \sigma\left(z^{(4)}\right) = \begin{Bmatrix} 0.74 \\ 0.93 \end{Bmatrix}$$

SOLUTION

$$\delta^{(4)} = \begin{Bmatrix} \dfrac{\left(a_1^{(4)} - y_1\right)}{2} \\ \dfrac{\left(a_2^{(4)} - y_2\right)}{2} \end{Bmatrix} = \begin{Bmatrix} -2.13 \\ -4.53 \end{Bmatrix}$$

$$\sigma'\left(z^{(4)}\right) = \sigma'\left(\begin{Bmatrix} 1.05 \\ 2.61 \end{Bmatrix}\right) = \begin{Bmatrix} \sigma'(1.05) \\ \sigma'(2.61) \end{Bmatrix} = \begin{Bmatrix} \sigma(1.05)(1 - \sigma(1.05)) \\ \sigma(2.61)(1 - \sigma(2.61)) \end{Bmatrix} = \begin{Bmatrix} 0.19 \\ 0.06 \end{Bmatrix}$$

$$\frac{\partial C}{\partial b^{(3)}} = \delta^{(4)} \odot \sigma'\left(z^{(4)}\right) = \begin{Bmatrix} -2.13 \times 0.19 \\ -4.53 \times 0.06 \end{Bmatrix} = \begin{Bmatrix} -0.41 \\ -0.29 \end{Bmatrix}$$

$$\frac{\partial C}{\partial w^{(3)}} = \frac{\partial C}{\partial b^{(3)}} {a^{(3)}}^{T} = \begin{bmatrix} -0.41 \\ -0.29 \end{bmatrix} \begin{bmatrix} 0.72 & 0.92 & 0.98 & 1 \end{bmatrix} = \begin{bmatrix} -0.3 & -0.38 & -0.4 & -0.41 \\ -0.21 & -0.27 & -0.29 & -0.29 \end{bmatrix}$$

$$\delta^{(3)} = {w^{(3)}}^{T} \frac{\partial C}{\partial b^{(3)}} = \begin{bmatrix} 0.1 & 0.5 \\ 0.2 & 0.6 \\ 0.3 & 0.7 \\ 0.4 & 0.8 \end{bmatrix} \begin{bmatrix} -0.41 \\ -0.29 \end{bmatrix} = \begin{Bmatrix} -0.19 \\ -0.26 \\ -0.33 \\ -0.4 \end{Bmatrix}$$

SOLUTION

$$\sigma'\left(z^{(3)}\right) = \begin{Bmatrix} 0.2 \\ 0.07 \\ 0.02 \\ 0 \end{Bmatrix}$$

$$\frac{\partial C}{\partial b^{(2)}} = \delta^{(3)} \odot \sigma'\left(z^{(3)}\right) = \begin{Bmatrix} -0.19 \times 0.2 \\ -0.26 \times 0.07 \\ -0.33 \times 0.02 \\ -0.4 \times 0 \end{Bmatrix} = \begin{Bmatrix} -0.04 \\ -0.02 \\ -0.01 \\ 0 \end{Bmatrix}$$

$$\frac{\partial C}{\partial w^{(2)}} = \frac{\partial C}{\partial b^{(2)}} {a^{(2)}}^T = \begin{bmatrix} -0.04 \\ -0.02 \\ -0.01 \\ 0 \end{bmatrix} \begin{bmatrix} 0.69 & 0.86 & 0.94 & 0.98 \end{bmatrix} = \begin{bmatrix} -0.03 & -0.03 & -0.03 & -0.04 \\ -0.01 & -0.02 & -0.02 & -0.02 \\ 0 & 0 & -0.01 & -0.01 \\ 0 & 0 & 0 & 0 \end{bmatrix}$$

$$\delta^{(2)} = {w^{(2)}}^T \frac{\partial C}{\partial b^{(2)}} = \begin{bmatrix} 0.1 & 0.5 & 0.9 & 1.3 \\ 0.2 & 0.6 & 1 & 1.4 \\ 0.3 & 0.7 & 1.1 & 1.5 \\ 0.4 & 0.8 & 1.2 & 1.6 \end{bmatrix} \begin{bmatrix} -0.04 \\ -0.02 \\ -0.01 \\ 0 \end{bmatrix} = \begin{Bmatrix} -0.02 \\ -0.03 \\ -0.03 \\ -0.04 \end{Bmatrix}$$

SOLUTION

$$\sigma'\left(z^{(2)}\right) = \begin{Bmatrix} 0.21 \\ 0.12 \\ 0.05 \\ 0.02 \end{Bmatrix}$$

$$\frac{\partial C}{\partial b^{(1)}} = \delta^{(2)} \odot \sigma'\left(z^{(2)}\right) = \begin{Bmatrix} -0.02 \times 0.21 \\ -0.03 \times 0.12 \\ -0.03 \times 0.05 \\ -0.04 \times 0.02 \end{Bmatrix} = \begin{Bmatrix} -0.004 \\ -0.003 \\ -0.002 \\ -0.001 \end{Bmatrix}$$

$$\frac{\partial C}{\partial w^{(1)}} = \frac{\partial C}{\partial b^{(1)}} {a^{(1)}}^T = \begin{bmatrix} -0.004 \\ -0.003 \\ -0.002 \\ -0.001 \end{bmatrix} \begin{bmatrix} 0.5 & 1 & 1.5 \end{bmatrix} = \begin{bmatrix} -0.002 & -0.004 & -0.006 \\ -0.002 & -0.003 & -0.005 \\ -0.001 & -0.002 & -0.003 \\ 0 & -0.001 & -0.001 \end{bmatrix}$$